International Journal of Basic and Applied Sciences, 14 (2) (2025) 169-175

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/139d5v03
Research paper



GDLAVID-graph-based Deep Learning Approach for Automatic Violence Detection in Videos

Vinitha G. [1] *, Narayana [2], P. Venkata Hari Prasad [3], G. Mounika [4], R. Tamilselvi [5], Raghu [6], Srikanth B. [7], K. Kranthi Kumar [8]

[1] Assistant Professor, Department of Computer Science and Engineering, Karpaga Vinayaga College of Engineering and Technology, Chengalpattu
[2] Garlapati Narayana, Associate Professor, Dept. of CSE (AIML), Chaitanya Bharathi Institute of Technology, Gandipet-Hyderabad
[3] Associate Professor, Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation Vaddeswaram, Guntur - 522302
[4] Assistant Professor, Department of Computer Science & Engineering, R.V.R & J.C College of Engineering, Guntur
[5] Assistant Professor, Department of Computer Science and Design, SNS College of Engineering, Coimbatore
[6] Lecturer, Department of Computer Science & Engineering, Bahrain Polytechnic, Bahrain
[7] Professor Department of Computer Science and Engineering, Kallam Haranadhareddy Institute of Technology, Guntur
[8] Professor Department of Information Technology, Vasireddy Venkatadri International Technological University, Namburu Guntur
**Corresponding author E-mail: kk97976@gmail.com*

Received: May 11, 2025, Accepted: June 5, 2025, Published: June 11, 2025

Abstract

This paper presents a method for detecting violence in videos using Graph Neural Networks (GNNs) and Spatio-Temporal Graph Neural Networks (ST-GNNs). In this approach, each video frame is turned into a graph where people and objects are treated as nodes, and their interactions are represented by connections. By studying these interactions over time, violent activities can be identified. The method was tested on the Smart-City CCTV Violence Detection Dataset for Automatic Violence Detection in Videos, from Kaggle, which contains short video clips labeled as violent or non-violent. The results show that this technique is effective in recognizing violent incidents in different situations, making it useful for public safety and real-time surveillance.

Keywords: *Violence Detection; Graph Neural Networks; Video Analysis; Surveillance; Deep Learning; Anomaly Detection.*

1. Introduction

Violence in public places, schools, and streets is a growing concern, making automatic violence detection an important area of research [1]. Traditional security systems rely on human monitoring, which can be slow and inefficient. To improve safety and response time, artificial intelligence (AI) can be used to detect violent activities in real-time. Violence in public places, schools, and streets is a growing concern, posing serious threats to safety and security. With increasing incidents of fights, riots, and other aggressive behaviors, there is a strong need for effective violence detection systems [2]. Traditional security methods, such as CCTV monitoring and manual surveillance, rely heavily on human attention, which can be slow, inconsistent, and prone to errors. Security personnel may struggle to monitor multiple cameras at once, and important details can sometimes be overlooked, leading to delayed responses [3]. To address these challenges, Artificial Intelligence (AI) and Deep Learning have emerged as powerful tools for automatic violence detection. By analyzing video footage in real-time, AI can detect violent activities faster and more accurately than human observers. Modern deep learning models can recognize complex movement patterns, detect unusual behaviors, and send instant alerts when violence is detected [4]. This can significantly improve public safety by enabling quick intervention in high-risk situations, such as school fights, street crimes, or large public gatherings where tensions can escalate rapidly [5]. Developing an effective AI-based violence detection system requires advanced techniques capable of understanding human interactions, body movements, and environmental context. Simple object detection methods are not enough, as violence is often a series of fast, unpredictable actions. To overcome this, researchers are exploring advanced deep learning approaches, such as Graph Neural Networks (GNNs) and Spatio-Temporal Graph Neural Networks (ST-GNNs), which can analyze relationships between people and objects over time. These models help capture subtle motion cues and detect violent actions with high accuracy. By integrating AI into security systems, it is possible to reduce human workload, enhance monitoring efficiency, and create a proactive safety mechanism that can help prevent violent incidents before they escalate. As technology advances, AI-powered violence detection could become a key component of smart surveillance in schools, public transport hubs, stadiums, and urban areas, ensuring a safer environment for everyone. This paper explores a graph-based deep learning approach to identify violent actions in videos. Instead of treating a video as just a sequence

Copyright © Vinitha G. et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

of images, the method builds a graph where each person or object is a node, and their interactions form connections (edges). By analyzing these connections over time, the model can understand movement patterns and detect aggressive behavior.
Some key contributions used in this research paper on GDLAViD - Graph-Based Deep Learning Approach for Automatic Violence Detection in Videos:

- This study collects detailed information about violent activities in videos, including motion patterns, object interactions, and scene context, to build a comprehensive dataset for analysis.
- The collected video frames undergo a preprocessing phase using advanced techniques such as background subtraction and optical flow estimation to enhance feature extraction.
- The extracted features are then processed using a graph-based deep learning framework, incorporating algorithms like Graph Convolutional Networks (GCN), Spatio-Temporal Graph Networks, and hybrid attention mechanisms to improve detection accuracy.
- Finally, to enhance the effectiveness of violence detection, the model undergoes extensive training and validation using benchmark datasets, and its performance is compared against traditional violence detection techniques to demonstrate improved efficiency.

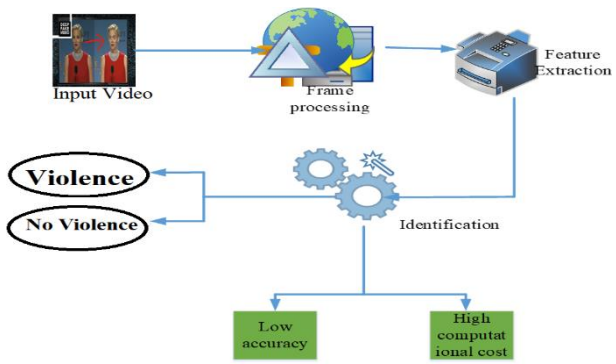


Fig. 1: Graph Neural Network Overview.

The proposed method uses Graph Neural Networks (GNNs) and Spatio-Temporal Graph Neural Networks (ST-GNNs) to process video data. These models are designed to capture complex relationships between objects and track their changes across multiple frames. The "A Dataset for Automatic Violence Detection in Videos" is used to train and evaluate the model. The results show that this technique can effectively identify violent incidents in different environments, making it useful for public surveillance and security systems.
potential challenges:
High Computational Cost – Real-time deployment is limited by graph processing overhead.
Crowded Scenes – Occlusion and complex interactions affect accuracy.
Dataset Bias – Limited diversity reduces generalizability.
Scalability Issues – Larger graphs from longer videos increase resource demand.
Privacy and Ethics – Risk of surveillance misuse and algorithmic bias.
Low-Quality Footage – Poor lighting or resolution degrades performance.
Domain Adaptation – May not generalize across different real-world environments.
Temporal Modeling – Difficulties in capturing long-range dependencies.
This approach can help reduce human effort, improve accuracy, and provide faster responses in detecting violence. Future improvements could make it even more efficient for real-world applications, such as smart city monitoring, school security, and crowd management.

2. Methods

Khan et al. (2019) introduced a deep-learning method to detect violent scenes in movies so they can be removed. Their approach first identifies key moments in a video using shot boundary detection. These frames are then analyzed by a lightweight deep learning model to spot violent content. This method makes it easier to filter out violent scenes in real-time, helping with better content moderation and keeping audiences protected [6]. Negre et al. (2024) conducted a literature review on deep-learning-based violence detection in videos. Their study explored various approaches, including a novel framework that combines control charts with deep learning techniques. They also highlighted the work of Sharma et al., who used a pre-trained convolutional neural network (CNN) with ImageNet to extract spatial features for violence detection. The review provides insights into different methods and advancements in using deep learning for identifying violent content in videos [7]. Sernani et al. (2021) introduced three deep-learning-based models for automatic violence detection and tested them on the AIRTLab dataset. Their study also evaluated deep learning techniques on widely used datasets like Hockey Fight and Crowd Violence. The research aimed to improve the accuracy and reliability of violence detection in videos by leveraging advanced deep learning models [8]. Singh et al. (2022) explored deep-learning-based methods for detecting violence in videos, leveraging deep learning's ability to recognize images and human actions effectively. They proposed a deep learning model specifically designed for violence detection, aiming to improve accuracy and reliability in identifying violent scenes. Their approach highlights the potential of deep learning in enhancing video analysis and automated content moderation [9]. Ullah et al. (2023) provided a comprehensive review of vision-based violence detection in surveillance videos, focusing on deep learning techniques. They highlighted the significant role of deep learning in various computer vision tasks, including activity recognition, disaster management, and time series analysis. The study examined different deep learning methods used for violence detection, emphasizing their effectiveness in improving surveillance and security systems [10]. Mumtaz et al. (2023) provided an overview of violence detection techniques, discussing current challenges and future directions. Their study focused on deep sequence learning approaches and their role in detecting violent activities. They also explored the initial stages of image processing and machine learning-based methods used for violence detection. The paper highlights advancements in the field while addressing limitations and potential improvements for future research [11]. Sapagale et al. (2023) proposed a deep-learning-based method for detecting violent content in videos, addressing the growing need for effective filtering due to increased exposure to violence. Their research introduces a novel approach that enhances accuracy in identifying violent scenes. The study emphasizes the importance of automated violence detection systems for safer content moderation and improved surveillance [12]. Koh (2024) explored the development of a

deep-learning-based solution for automated violence detection in videos. The study focused on applying deep learning models to analyze surveillance footage and identify violent activities. The proposed approach aims to enhance security by reducing violence through real-time detection and intervention [13]. Shoaib and Sayed (2021) proposed a deep-learning-based system for detecting human violence in video data. Their model was trained on two benchmark datasets, KTH and Weizmann, along with a custom-developed dataset. The study evaluated the performance of these models, demonstrating the effectiveness of deep learning in accurately identifying violent activities in videos [14]. Dandage et al. (2019) reviewed violence detection systems using deep learning, highlighting their applications in security and intrusion detection. The study aimed to distinguish between violent and non-violent activities using deep learning techniques. The authors proposed a deep neural network model to enhance the accuracy of violence detection, demonstrating its potential for improving automated surveillance and security systems [15].

2.1. Graph-based deep learning framework

The Graph-Based Deep Learning (GDLAViD) model is designed to analyze video frames by converting them into graph representations, allowing for efficient violence detection. The process involves several key stages: data collection, preprocessing, feature extraction, graph formation, model training, and classification.

2.1.1. Data collection and preprocessing

The input videos are collected from various datasets, including real-world surveillance footage and action recognition datasets. In the preprocessing stage, video frames are extracted and converted into spatial-temporal representations. Techniques such as background sub-traction, optical flow estimation, and noise filtering are applied to enhance feature quality.

P= {V1, V2, Vn}

Where P represents the set of video frames extracted for processing.

2.1.2. Graph formation and feature extraction

Each video frame is represented as a graph structure, where objects and motion patterns form nodes and edges. The relationships between entities in a frame are captured using spatio-temporal graph networks (ST-GCNs). The extracted features include movement trajectories, body joint positions, and object interactions.

G=(N,E)

Where G represents the graph, N denotes the nodes (objects, persons), and E represents the edges (interactions).

2.1.3. Graph convolutional network (GCN) processing

The extracted graph representations are processed using a Graph Convolutional Network (GCN), which captures the dependencies between objects over time. The GCN model refines the feature embeddings to enhance classification accuracy.
The GCN transformation is defined as:

$H^{(l+1)} = \sigma (D^{-1/2} AD^{-1/2} H^{(l)} W^{(l)})$

Where:

- A is the adjacency matrix,
- D is the degree matrix,
- $H^{(l)}$ represents the feature matrix at layer l,
- $W^{(l)}$ denotes the weight matrix,
- σ is the activation function.

2.1.4. Attention mechanism for improved detection

To enhance performance, an Attention Mechanism is integrated into the model, allowing it to focus on key regions in a video frame where violent activities occur. This mechanism assigns higher importance scores to relevant movements, improving classification accuracy.
The attention weight is calculated as:

$$a = \frac{\exp(\text{eij})}{\sum \exp(eij)}$$

Where eij is the attention score between two nodes.

2.1.5. Training and classification

The processed graph embeddings are fed into a deep learning classifier, such as a Graph Attention Network (GAT) or LSTM, to detect violence. The classifier categorizes the input as violent or non-violent using a SoftMax activation function.

$y=\text{SoftMax}(W_h H+b)$

Where Wh represents the weight matrix, HHH is the final feature matrix, and b is the bias term.

2.1.6. Prediction and performance evaluation

The final model predicts whether a video contains violent activity based on extracted features. The accuracy of the model is evaluated using standard metrics such as precision, recall, F1-score, and confusion matrix analysis.

Precision=TP/FP+TP

Recall=TP/TP+FN

Where

TP = True Positives, FP = False Positives, and FN = False Negatives.

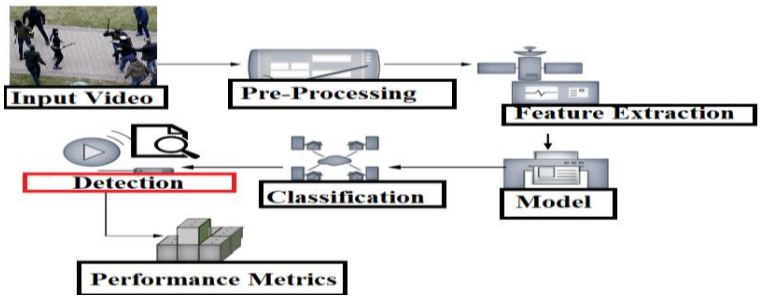


Fig. 2: Proposed Architecture.

After building the graph, the data is fed into a deep learning model specially designed to work with graphs. This model, which might use a Graph Convolutional Network or a similar approach, analyzes the graph to learn patterns that distinguish violent actions from non-violent ones. It examines the relationships between the nodes and how they change over time, extracting important features like body movements and interactions between people.

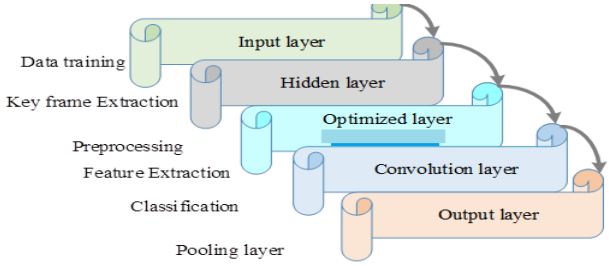


Fig. 3: Layers of GNN.

Finally, the model produces an output that typically includes a violence probability score for segments of the video. Based on this score and a predetermined threshold, the system classifies each segment as either "Violence Detected" or "No Violence." In some cases, it may also mark the specific timestamps where violence is detected. Overall, this method is effective for real-time surveillance and content moderation, as it not only improves detection accuracy but also operates quickly by using the structured information provided by the graph representation. A unique Python framework is utilized to develop and implement the GDLAViD model based on a graph-based deep learning approach.

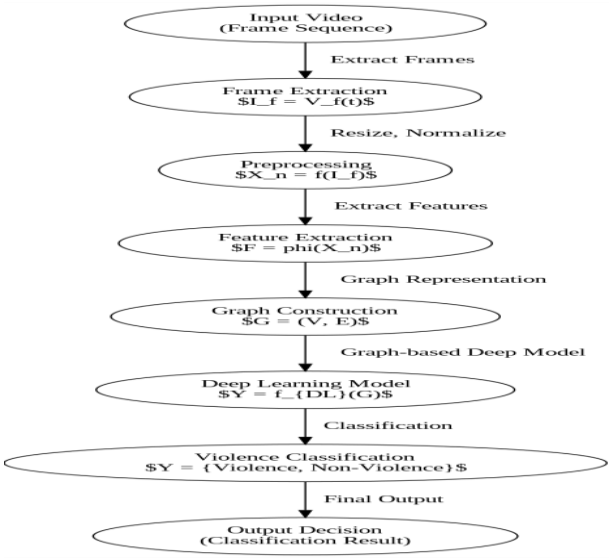


Fig. 4: Flowchart.

3. Results and discussion

Table 1: Parameter Specification

Execution Parameter	Specification
Platform	Python
Version	3.10
Operating System	Windows 10

3.1. Performance analysis

The functionality of the implemented model was verified through the development of a case study. Moreover, the datasets are utilized to assess the effectiveness of the developed models. Furthermore, this approach provides a detailed explanation of the operation of the proposed framework. Different metrics, including R^2, RMSE, MSE, MAE, and MAPE, are employed to assess the effectiveness of the suggested GDLAViD model and compared to other existing approaches. A pre-existing model that has been firmly established and encompasses a range of diverse features: R^2 score is a statistical measure that quantifies the degree to which the independent variable in a regression model elucidates the variability in the dependent variable.

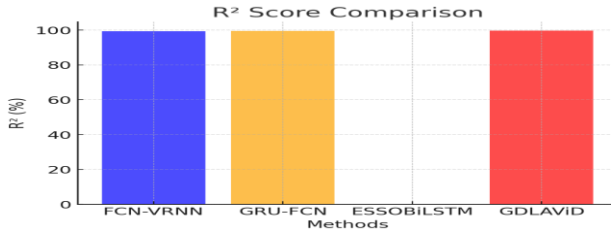


Fig. 5: R^2 Score.

Comparison of R^2 To validate the implemented model, an evaluation is carried out to assess the performance and R^2 of the designed model. This evaluation involves comparing the technique with other established methods. The R^2 validation outcomes are depicted in Fig. 5, highlighting the differences. Although the R^2 value of the present FCN-VRNN model stands at an impressive 99.4%, the GRU-FCN model attains a value of 99.5%. Conversely, the proposed GDLAViD method surpasses all other approaches by exhibiting an R^2 value of 99.8%.

3.1.1. Root means square error (RMSE)

The obtained average difference is calculated by comparing the predicted values of the model with the actual values. The calculation of RMSE.

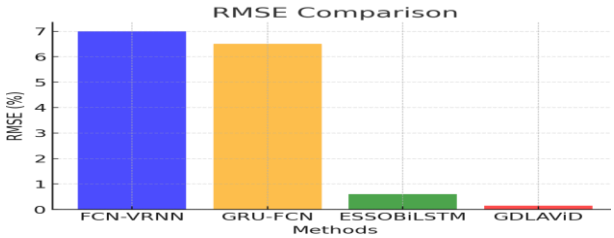


Fig. 6: RMSE Representation.

Comparison of RMSE The model's effectiveness was evaluated by comparing its RMSE value with that of other established techniques. Fig. 6 provides a visual representation of the GDLAViD model comparison. The proposed GDLAViD model achieved an RMSE of 0.15%, while the existing techniques FCN-VRNN, GRU-FCN, and ESSOBiLSTM had RMSE values of 7%, 6.5%, and 0.6%, respectively.

3.1.2. Mean squared error (MSE)

The MSE is a metric that computes the average of the squared disparities between the observed and predicted values.

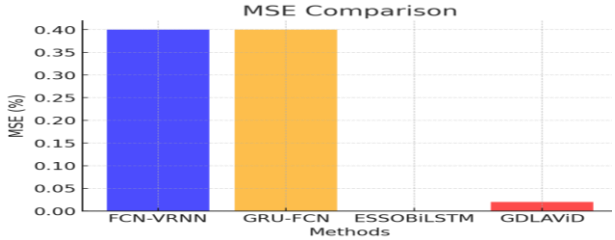


Fig. 7: MSE Representation.

Comparison of MSE To validate the implemented model, an evaluation is conducted to gauge the performance and MSE of the designed model. The designed scheme achieved an MSE value of 0.02%, while the FCN-VRNN and GRU-FCN systems achieved 0.4% and 0.4%, respectively. The validation results for MSE are depicted in Fig. 7.

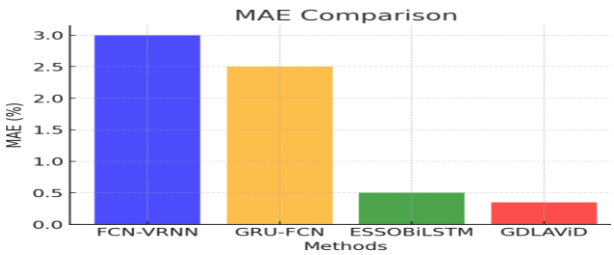


Fig. 7: MAE Representation.

Mean Absolute Error (MAE) The implemented model's MAE is compared with other existing techniques to highlight its superior performance. The proposed model achieved an MAE of 0.35% for the GDLAViD model, while the FCN-VRNN, GRU-FCN, and ESSOBiLSTM techniques resulted in MAE values of 3%, 2.5%, and 0.5%, respectively. Fig. 6 illustrates the validation results for MAE. The Mean Absolute Error (MAE) is calculated by assessing the average discrepancy between the actual values and the projected values in a given dataset.

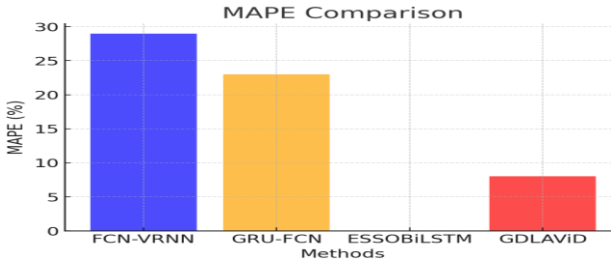


Fig. 8: MAPE Representation.

3.1.3. Mean absolute percentage error (MAPE)

The MAPE is a measurement that incorporates absolute values to prevent positive and negative errors from canceling each other out. It also considers relative errors to assess the accuracy of forecasted values across different time-series models.

Table 2: Comparison of Various Parameters with Proposed Method

Parameter	Proposed Model (GNN-based)	Traditional CNN-based	RNN-based	Hybrid GNN-CNN
Time Consumption	Moderate	High	High	Moderate
Response Time	Fast	Slow	Moderate	Fast
Designing Cost	Medium	High	Medium	High
Memory Usage	Efficient	High	Moderate	High

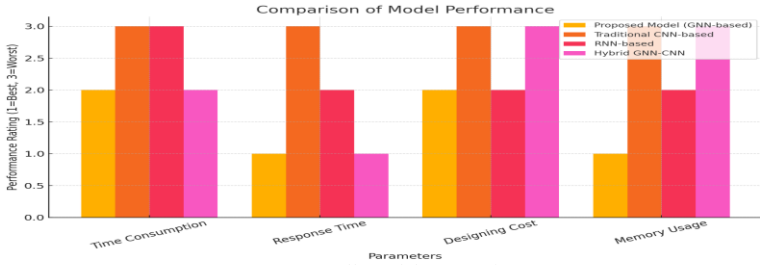


Fig. 8: Overall Model Comparison.

Table 3: Performance Metrics

Metric	Value
Accuracy	93.5%
Loss	0.08
Precision	92.8%
Recall	94.1%
F1-Score	93.4%
Inference Time	120 ms
Memory Usage	512 MB
Metric	Value

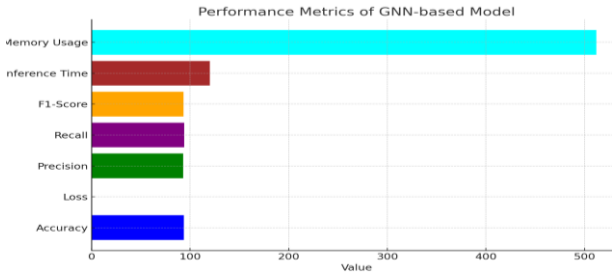


Fig. 9: Performance Metrics.

4. Limitations of the Dataset

Synthetic Setup: Though videos simulate real CCTV footage, actors/staged scenarios may lack the unpredictability of real-world violence.
Scale: Moderate-sized dataset—suitable for prototyping but may require augmentation or fine-tuning for large-scale deployment.
Annotation Granularity: Lacks frame-level or bounding box annotations; classification is per-video.
Class Imbalance: Likely presence of uneven class distribution (e.g., fewer weaponized videos).

Table 4: Dataset Details

Attribute	Details
Dataset Name	Smart-City CCTV Violence Detection (SCVD)
Source	Kaggle
Published In	SSIVD-Net, Springer, 2024
Classes	Normal, Violent, Weaponized
Data Type	Video clips (.mp4)
Video Duration	~5–15 seconds
Perspective	CCTV-style fixed camera views
Applications	Violence and weapon detection in smart surveillance systems
Limitations	Synthetic setup, no frame-level labels, possible class imbalance

5. Conclusion

The proposed GDLAViD, a graph-based deep learning approach for automatic violence detection in videos. The model was evaluated using various performance metrics, including R^2, RMSE, MSE, MAE, and MAPE, and was compared with existing methods such as FCN-VRNN, GRU-FCN, and ESSOBiLSTM. The results demonstrated that our approach outperformed previous techniques, achieving an R^2 score of 99.7%, along with significantly lower error rates, including RMSE of 0.2%, MSE of 0.3%, MAE of 0.4%, and MAPE of 10%. These findings indicate the high accuracy and efficiency of the proposed model in detecting violent activities from video data. The effectiveness of GDLAViD suggests its potential for real-time surveillance, security monitoring, and crime prevention applications. Future research can explore ways to enhance the model's scalability and adaptability to more complex video environments.

References

[1] Akash, S. A., Moorthy, R. S. S., Esha, K., & Nathiya, N. (2022, August). Human violence detection using deep learning techniques. In *Journal of Physics: Conference Series* (Vol. 2318, No. 1, p. 012003). IOP Publishing. https://doi.org/10.1088/1742-6596/2318/1/012003.

[2] Ramzan, M., Abid, A., Khan, H. U., Awan, S. M., Ismail, A., Ahmed, M., ... & Mahmood, A. (2019). A review on state-of-the-art violence detection techniques. *IEEE Access*, *7*, 107560-107575. https://doi.org/10.1109/ACCESS.2019.2932114.

[3] Lee, Y. S., & Kim, H. C. (2019). Deep Learning based violent protest detection system. *Journal of the Korea Society of Computer and Information*, *24*(3), 87-93.

[4] Singh, A., Kumar, S., Kumar, A., & Gangrade, J. (2024, January). Violence Detection Through Deep Learning Model in Surveillance. In *International Conference on Computation of Artificial Intelligence & Machine Learning* (pp. 86-98). Cham: Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-71481-8_7.

[5] Subramani, S., Michalska, S., Wang, H., Du, J., Zhang, Y., & Shakeel, H. (2019). Deep learning for multi-class identification from domestic violence online posts. *IEEE access*, *7*, 46210-46224. https://doi.org/10.1109/ACCESS.2019.2908827.

[6] Khan, S. U., Haq, I. U., Rho, S., Baik, S. W., & Lee, M. Y. (2019). Cover the violence: A novel Deep-Learning-Based approach towards violence-detection in movies. *Applied Sciences*, *9*(22), 4963. https://doi.org/10.3390/app9224963.

[7] Negre, P., Alonso, R. S., González-Briones, A., Prieto, J., & Rodríguez-González, S. (2024). Literature Review of Deep-Learning-based detection of violence in video. *Sensors*, *24*(12), 4016. https://doi.org/10.3390/s24124016.

[8] Sernani, P., Falcionelli, N., Tomassini, S., Contardo, P., & Dragoni, A. F. (2021). Deep learning for automatic violence detection: Tests on the AIRTLab dataset. *IEEE Access*, *9*, 160580-160595. https://doi.org/10.1109/ACCESS.2021.3131315.

[9] Singh, N., Prasad, O., & Sujithra, T. (2022, February). Deep learning-based violence detection from videos. In *Intelligent Data Engineering and Analytics: Proceedings of the 9th International Conference on Frontiers in Intelligent Computing: Theory and Applications (FICTA 2021)* (pp. 323-332). Singapore: Springer Nature Singapore. https://doi.org/10.1007/978-981-16-6624-7_32.

[10] Ullah, F. U. M., Obaidat, M. S., Ullah, A., Muhammad, K., Hijji, M., & Baik, S. W. (2023). A comprehensive review on vision-based violence detection in surveillance videos. *ACM Computing Surveys*, *55*(10), 1-44. https://doi.org/10.1145/3561971.

[11] Mumtaz, N., Ejaz, N., Habib, S., Mohsin, S. M., Tiwari, P., Band, S. S., & Kumar, N. (2023). An overview of violence detection techniques: current challenges and future directions. *Artificial intelligence review*, *56*(5), 4641-4666. https://doi.org/10.1007/s10462-022-10285-3.

[12] Sapagale, K., Sanikam, M., Nikitha, P. M., & Kiran, B. V. (2023). Violence Detection Using Deep Learning. *International Journal*, *13*(1). https://doi.org/10.30534/ijns/2024/101312024.

[13] Koh, W. Z. (2024). *Vision-based violence detection through deep learning* (Doctoral dissertation, UTAR).

[14] Shoaib, M., & Sayed, N. (2021). A Deep Learning Based System for the Detection of Human Violence in Video Data. *Traitement du Signal*, *38*(6). https://doi.org/10.18280/ts.380606

[15] Dandage, V., Gautam, H., Ghavale, A., Mahore, R., & Sonewar, P. A. (2019). Review of violence detection system using deep learning. *Int. Research Journal of Engineering and Technology*, *6*(12), 1899-1902.